REMARKS

An Excess Claim Fee Payment letter is submitted herewith to cover the cost of four (4) additional total claims.

Claims 1-12 and 14-28 are all the claims presently pending in the application. Claims 1, 9, 14, 17, 21 and 23 have been amended to more particularly define the claimed invention. Claims 25-28 have been added to claim additional features of the claimed invention.

It is noted that the claim amendments are made only for more particularly pointing out the invention, and not for distinguishing the invention over the prior art, narrowing the claims or for any statutory requirements of patentability. Further, Applicant specifically states that no amendment to any claim herein should be construed as a disclaimer of any interest in or right to an equivalent of any element or feature of the amended claim.

Claims 1-12 and 14-24 stand rejected under 35 U.S.C. § 103(a) as being allegedly unpatentable over Lewak et al. (U. S. Patent No. 5,544,360) in view of Goldszmidt et al. ("A Probabilistic Approach to Full-Text Document Clustering" 1998, Technical Report ITAD-433-MS-98-044, SRI International).

This rejection is respectfully traversed in the following discussion.

I. THE CLAIMED INVENTION

The claimed invention (e.g., as recited in claim 1) is directed to a method (e.g., a computer-implemented method) for identifying relationships between text documents and structured variables pertaining to the text documents. The method includes providing a dictionary of keywords in the text documents, forming categories of the text documents using the dictionary and an automated algorithm, counting occurrences of the structured variables, the categories, and combinations of the structured variables and the categories for the text documents, and calculating probabilities of occurrences of the combinations of structured variables and categories.

Importantly, the method includes identifying a relationship between a structured variable (of the structured variables) and text documents included in a category (of the categories) based on a probability of occurrence of a combination of the structured variable and the category (Application at page 11, line 10-page 12, line 7; page 16, line 11-page 17, line 11; Figures 11 and 14; page 18, line 10-page 20, line 14).

Conventional methods of analyzing text documents cannot efficiently (e.g., automatically) identify interesting relationships between text documents (e.g., unstructured

free-form text documents) and structured variables. Instead, words and phrases which frequently occur in the documents are plotted on a graph and users are required to determine for themselves whether an interesting relationship exists, which is labor intensive and time consuming (Application at page 1, line 17-page 2, line 1).

The claimed invention, on the other hand, identifies a relationship between a structured variable (of the structured variables) and text documents included in a category (of the categories) based on a probability of occurrence of a combination of the structured variable and the category (Application at page 11, line 10-page 12, line 7; page 16, line 11-page 17, line 11; Figures 11 and 14; page 18, line 10-page 20, line 14). Thus, unlike conventional methods, the claimed invention can efficiently (e.g., automatically) identify interesting relationships between the structured variables and categories of text documents (Application at page 11, lines 10-11; page 23, lines 1-8).

II. THE ALLEGED PRIOR ART REFERENCES

The Examiner alleges Lewak would have been combined with Goldszmidt to form the invention of claims 1-12 and 14-24. Applicant submits, however, that these references would not have been combined and even if combined, the combination would not teach or suggest each and every element of the claimed invention.

Lewak discloses a method for accessing computer files. In the Lewak method, the user defines hybrid folders by describing the file contents of the files that belong to particular hybrid folders (Lewak at Abstract).

Goldszmidt discloses a probabilistic approach to full-text document clustering which includes scoring document similarity based on probabilistic considerations. Similarity is scored according to the expectation of the same words appearing in two documents. The score enables the investigation of different smoothing methods for estimating the probability of a word appearing in a document, for purposes of clustering (Goldszmidt at Abstract).

Applicant respectfully submits that these references would not have been combined as alleged by the Examiner. Indeed, Lewak is directed to a method for accessing computer files, whereas Goldszmidt is directed to a method which estimates the probability of a word appearing in a document, for purposes of document clustering. Thus, these references are completely unrelated, and no person of ordinary skill in the art would have considered combining these disparate references, absent impermissible hindsight.

Further, these references clearly do not teach or suggest their combination. Therefore,

Applicant respectfully submits that one of ordinary skill in the art would not have been so motivated to combine the references as alleged by the Examiner. Therefore, the Examiner has failed to make a prima facie case of obviousness.

Moreover, neither Lewak, nor Goldszmidt, nor any alleged combination teaches or suggests a method for identifying relationships between text documents and structured variables pertaining to the text documents, which includes "*identifying a relationship between a structured variable of said structured variables and text documents included in a category of said categories based on a probability of occurrence of a combination of said structured variable and said category*", as recited, for example, in claims 1, 14, 17 and 23. As noted above, unlike conventional methods, the claimed invention can efficiently (e.g., automatically) identify interesting relationships between the structured variables and categories of text documents (Application at page 11, lines 10-11; page 23, lines 1-8).

Clearly, these novel features are not taught or suggested by the cited references or their combination. Indeed, the Examiner expressly concedes that Lewak does not teach or suggest this feature on page 4 of the Office Action.

The Examiner again alleges that the feature is taught by Goldszmidt. As Applicant stated in detail in the Amendments filed on April 26, 2005 and October 11, 2005, the Examiner is incorrect.

Indeed, it is important for the Examiner to understand that the claimed invention **identifies a relationship between A STRUCTURED VARIABLE (e.g., a time period) and text documents included in a category (e.g., all text documents including a particular keyword)**. The relationship is identified based on a probability of occurrence of a combination of the structured variable (e.g., the time period) and the category (e.g., all text documents including a particular keyword).

For example, an exemplary embodiment of the claimed invention is described in the present Application at Figures 11 and 14 and page 18, line 10-page 20, line 14. In this exemplary embodiment, the claimed invention counted 15 occurrences of the combination of July 1998 (e.g., a structured variable) and text documents including the keyword "elimination" (e.g., a category of text documents) (Application at page 20, lines 2-5; Figure 14). This exemplary embodiment of the claimed invention may, therefore, identify a relationship between July 1998 (e.g., a structured variable) and text documents including the keyword "elimination" based on a probability of occurrence of this combination.

Goldszmidt has absolutely nothing to do with this novel feature.

Indeed, Applicant would point out that the Examiner again does not even allege that Goldsmidt has anything to do with a structured variable. Instead, the Examiner merely states that "Goldszmidt teaches a similarity measure based on probability (an overlap measure) which measures the degree of overlap between pairs of documents" (Office Action at page 5). That is, the Examiner again attempts to equate "measuring the degree of overlap between pairs of documents" (as in Goldszmidt) with identifying a relationship between text documents and structured variables (e.g., structured data such as a day, month year, etc.). Clearly, this is completely unreasonable.

Applicant would point out that in spite of his request that the Examiner provide some support for his position that merely identifying clusters of "documents" would be equated by one of ordinary skill in the art identifying a relationship between text documents and structured variables (e.g., structured data), the Examiner has provided no such support but merely continues to repeat the same general assertions.

Again, Applicant would point out that Goldszmidt merely teaches a **document similarity** metric. Clearly, this document similarity metric does not teach or suggest the claimed invention. Indeed, the claimed invention is not necessarily intended to measure similarity **between documents**. Instead, an exemplary aspect of the claimed invention may involve looking for a correlation between some event in the document text (e.g., the occurrence of a text category or word) and some other structured variable (e.g., date of the document, author's country of origin, etc.).

The probability measure in the claimed invention is simply a manner in which the claimed invention may determine whether any relationship (e.g., correlation) might exist between two supposedly independent variables (text and structure) and if so, what that correspondence is (e.g. certain words or phrases used more often at certain times than at other times).

As noted above, in complete contrast to the claimed invention, Goldszmidt merely teaches a document similarity metric and does not teach or suggest a method of identifying relationships **between text documents and structured variables.**

Therefore, even assuming (arguendo) that these references would have been combined, and even assuming (arguendo) that the Examiner's allegations regarding the teachings of Goldszmidt are correct, the combination of Lewak and Goldszmidt clearly does

not teach or suggest the claimed invention.

Therefore, Applicant submits that these references would not have been combined and even if combined, the combination would not teach or suggest each and every element of the claimed invention. Therefore, the Examiner is respectfully requested to withdraw this rejection.

III. FORMAL MATTERS AND CONCLUSION

In view of the foregoing, Applicant submits that claims 1-12 and 14-28, all the claims presently pending in the application, are patentably distinct over the prior art of record and are in condition for allowance. The Examiner is respectfully requested to pass the above application to issue at the earliest possible time.

Should the Examiner find the application to be other than in condition for allowance, the Examiner is requested to contact the undersigned at the local telephone number listed below to discuss any other changes deemed necessary in a telephonic or personal interview.

The Commissioner is hereby authorized to charge any deficiency in fees or to credit any overpayment in fees to Assignee's Deposit Account No. 09-0441.

Respectfully Submitted,

Date: 2/21/06

Phillip E. Miller, Esq.
Registration No. 46,060

McGinn IP Law Group, PLLC
8321 Old Courthouse Road, Suite 200
Vienna, VA 22182-3817
(703) 761-4100
Customer No. 21254

CERTIFICATE OF FACSIMILE TRANSMISSION

I hereby certify that the foregoing was filed by facsimile with the United States Patent and Trademark Office, Examiner James Blackwell, Group Art Unit #2176 at fax number 571-273-8300 this 21st day of February, 2006.

Phillip E. Miller
Reg. No. 46,060